Manuscript submitted to *The Econometrics Journal*, pp. 1–19.

Using post-regularization distribution regression to measure the effects of a minimum wage on hourly wages, hours worked and monthly earnings

Martin Biewen[†‡] and Pascal Erhardt[†]

[†] *University of Tübingen, School of Business and Economics, Mohlstr. 36, 72074 Tübingen, Germany.*
E-mail: martin.biewen@uni-tuebingen.de, pascal.erhardt@uni-tuebingen.de
[‡] *IZA Institute of Labor Economics, Schaumburg-Lippe-Straße 5-9, Bonn, Germany*

Summary We evaluate the distributional effects of a minimum wage introduction based on a data set with a moderate sample size but a large number of potential covariates. In this context, the selection of relevant control variables at each distributional threshold is crucial to test hypotheses about the impact of the continuous treatment variable. To this end, we use the post-double selection logistic distribution regression approach proposed by Belloni et al. (2018a), which allows for uniformly valid inference about the target coefficients of our low-dimensional treatment variables across the entire outcome distribution. Our empirical results show that the minimum wage replaced hourly wages below the minimum threshold, increased monthly earnings in the lower-middle segment but not at the very bottom of the distribution, and did not significantly affect the distribution of working hours.

Keywords: *wage structure, automatic specification search, double machine learning*

1. INTRODUCTION

The introduction of Germany's statutory minimum wage on January 1, 2015 was a significant policy experiment. While industry-specific minimum wages had existed before 2015, Germany was among the few countries worldwide without a universal minimum wage. The imposition of a nationwide minimum wage of 8.50 euros/hour in 2015 represented a major intervention in the German labour market, affecting around 4 million workers (more than 11% of the workforce) who earned less than 8.50 euros/hour before its introduction; see Mindestlohnkommission (2020).[1]

Based on different data sets, previous contributions have examined various aspects of the German minimum wage introduction. As to potential employment effects, the literature has reached the consensus that these were non-existent or very small; see, e.g. Caliendo et al. (2019), Dustmann et al. (2022), Bossler and Schank (2023) and Link (2024). By contrast, the literature appears to have reached conflicting results about the distributional effects of the minimum wage, i.e., its effects on the distributions of hourly wages, monthly earnings and working hours (the latter including potential shifts between full-time, part-time and marginal part-time work). Using register data, Bossler and Schank (2023) find that the minimum wage significantly reduced inequality in monthly wages. On the basis of survey data, however, Burauel et al. (2019, 2020) and Caliendo et al. (2022) conclude that the minimum wage introduction also reduced working hours, neutralizing its effect on monthly wages. Given that German register data do not include

[1] See Caliendo et al. (2019) for a more detailed overview of the institutional details of the minimum wage introduction.

information on working hours, Biewen et al. (2022) analyse large-scale data from the statistical offices to conclude that working hours were *not* causally affected by the minimum wage so that increased hourly wages should fully translate into changes in monthly earnings.

Given the inconclusive evidence, this paper reexamines the effects of the minimum wage on the distributions of hourly wages, monthly earnings, and working hours. We use the same survey data analysed in Burauel et al. (2019, 2020) and Caliendo et al. (2022). Based on modern machine learning methods that allow us to examine the effects of the minimum wage across all points of the distribution, we reach the conclusion that the minimum wage replaced hourly wages below the minimum threshold, increased monthly earnings in the lower-middle segment but not at the very bottom of the distribution (consistent with Bossler and Schank, 2023), and did not significantly affect the distribution of working hours. Our results help reconcile the conflicting results in the literature using different data sources mentioned above.

Our econometric analysis is based on the distribution regression approach introduced by Foresi and Peracchi (1995) and developed by Chernozhukov et al. (2013). Distribution regression involves performing numerous binary regressions, each modelling the probability that the outcome falls below a specific threshold, across a finely spaced grid covering the entire distribution. Compared to alternative methods such as conditional or unconditional quantile regression, distribution regression directly targets nominal points in the outcome distribution. It is therefore ideally suited to study changes in distributions, such as hourly wages, working hours or monthly wages, whose quantiles typically change over time, complicating the interpretation of quantile regression results if more than one time period is involved. Moreover, distribution regression easily deals with discrete mass points, see Chernozhukov et al. (2013), which is particularly relevant when dealing with discrete distributions (working hours) or distributions with severe heaping (hourly wages, especially after the introduction of a minimum wage). This is in contrast to conditional or unconditional quantile regression that are based on the assumption of continuous distributions.

A key challenge in distribution regressions is the need to specify and estimate multiple binary models. The different binary regression models should take into account potentially varying sets of covariates as different covariates may matter at different points of the distribution. This challenge is particularly pronounced when using a sample with a moderate sample size but a large number of potential covariates as it will become inevitable to select relevant covariates at each distributional threshold to save degrees of freedom. In light of the fact that this is generally required to be carried out across a substantial number of thresholds, a hand-picked approach might result in a considerable amount of arbitrary specification search with unknown consequences for the potential bias of estimated coefficients and their estimated standard errors. Another challenge is the likely high correlation between regression results at different thresholds, which makes inference across multiple points in the distribution more complex. Apart from inferential aspects, the sheer practical task of specification searches for a large number of parallel regressions suggests the use of machine learning techniques such as the lasso to separately predict nuisance terms at the large number of distributional thresholds.

Both the practical and the inferential aspect have been addressed by advances in econometric machine learning. In a recent contribution, Belloni et al. (2018a) showed how to employ a sequence of ℓ_1-regularized logistic regressions such that inferences about the coefficients of target regressors are valid both pointwise, i.e., in each regression model separately, as well as uniformly across a large number of models estimated. Their proposed algorithm is closely related to the concept of "partialling-out" in the econometrics literature, where one removes nuisance terms that are both related to the outcome and the treatment variables (in the present context referred to as "double selection"). Picking covariates by the lasso method also entails functional form specification as the features offered to the lasso may include arbitrary transformations of variables (logs, polynomials, indicators for particular values, interaction terms). The possibility to obtain valid inference after large-scale automatic specification search is a remarkable achievement of the recent econometric machine learning literature. It represents a major improvement over the often arbitrary and undocumented specification searches carried out by individual researchers, which are typically influenced in unknown ways by the propagation of pre-tested control variables used in previous research.

A key assumption of the ℓ_1-regularized methods used by us is approximate sparsity, i.e., the sequence of coefficients of potential confounder terms sorted by absolute value decays quickly enough, but does not have to be exactly equal to zero. This is a natural assumption in substantive applications where one would like to control for relevant confounding information, but does not rule out the existence of factors whose influence may be negligible for the inferential purpose at hand. In the following, we employ the method described in Belloni et al. (2018a), albeit with certain modifications to address the fact that our data and research design features observation clusters and sampling weights.[2]

2. ECONOMETRIC METHODS

We aim to measure the effects of the minimum wage introduction across the whole distribution of an outcome variable $Y \in \{$"hourly wage","hours worked","monthly earnings"$\}$. For this purpose, we use the logistic distribution regression model

$$\mathrm{P}(Y < u \mid D, X) = \mathrm{E}[\mathbf{1}\{Y < u\} \mid D, X] = \Lambda(D\theta_u + X\beta_u), \quad u \in \mathcal{U}. \tag{2.1}$$

As in Belloni et al. (2018a), $\Lambda(t) = \exp\{t\}/(1 + \exp\{t\}),\ t \in \mathbb{R}$ denotes the logistic (inverse) link function. This model measures the effects of target variables $D = (D_1, \ldots, D_{\tilde{p}})$, representing the difference-in-differences specification described below, across a set of grid points $u \in \mathcal{U}$ of the outcome distribution. In order to isolate the effect of the target variables on the likelihood of falling below a particular threshold u, it is necessary to control for confounders X, which may vary across different points of the outcome distribution. This motivates the use of a separate lasso procedure to select a set of relevant control variables at each point $u \in \mathcal{U}$.

For the rest of this article, let the vector of potential confounding features be given by $X = (X_1, \ldots, X_p)$, target coefficients $\theta_u = (\theta_{u1}, \ldots, \theta_{u\tilde{p}})'$, nuisance parameters $\beta_u = (\beta_{u1}, \ldots, \beta_{up})'$ as well as indicators $Y^u = \mathbf{1}\{Y < u\}$. Our application comprises clusters $g = 1, \ldots, G$ of observations $W_{ig} = (Y_{ig}, D_{ig}, X_{ig}) = (Y_{ig}, D_{1ig}, \ldots, D_{\tilde{p}ig}, X_{1ig}, \ldots, X_{pig})$, assuming that observations are independent across clusters but may be correlated within clusters; see next section for more details. The number of observations within a cluster is denoted by n_g such that $\sum_{g=1}^{G} n_g = n$. Furthermore, we make use of deterministic sampling weights v_{ig}, which are normalized to sum up to the total number of observations.

[2]Our paper appears to be one of the first substantive applications of the method described in Belloni et al. (2018a), apart from Chiang (2020) who also considers modifications for observation clusters.

2.1. Estimation

The post-double lasso method for the logistic regression model and other generalized linear models was developed by Belloni et al. (2016a). Belloni et al. (2018a) extended the method to cover uniform inference for functional parameters, e.g., for coefficients of many parallel logit models as needed in our application. Extensions of the lasso method to clustered data were considered in Belloni et al. (2016b), Chiang (2020) and Ahrens et al. (2020). Essentially, applying lasso to clustered data involves the treatment of blocks of observations belonging to the same cluster as “super-observations”, thereby “collapsing” these blocks and computing penalty levels based on partial sums of observations.

Following Belloni et al. (2018a) and applying modifications for clustering similar to Chiang (2020), as well as for sampling weights, we use the following post-double selection procedure to estimate the process of target coefficients $(\theta_{uj}), u \in \mathcal{U}, j \in \mathcal{J} = \{1, \ldots, \tilde{p}\}$.

Step 1. Post-lasso logit of Y^u on (D, X)

Run the ℓ_1-penalized logistic regression:

$$(\hat{\theta}_u, \hat{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, M_u(W_{ig}, \theta, \beta) + \frac{\lambda_1}{G} \|\hat{\Psi}_u(\theta', \beta')'\|_1, \tag{2.2}$$

$$M_u(W_{ig}, \theta, \beta) = \log(1 + \exp(D_{ig}\theta + X_{ig}\beta)) - Y_{ig}^u \cdot (D_{ig}\theta + X_{ig}\beta), \tag{2.3}$$

where λ_1 is a penalty parameter and $\hat{\Psi}_u$ a diagonal penalty loading matrix whose entries are chosen according to the procedure explained in Appendix A.1. Define $\hat{\mathcal{S}}_u = \text{supp}(\hat{\theta}_u', \hat{\beta}_u') = \{l \in \{1, \ldots, \tilde{p}, \ldots, \tilde{p} + p\} \mid (\hat{\theta}_u', \hat{\beta}_u')_l \neq 0\}$.

Obtain the post-ℓ_1 logit coefficients:

$$(\tilde{\theta}_u, \tilde{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, M_u(W_{ig}, \theta, \beta) : \text{supp}(\theta', \beta') \subseteq \hat{\mathcal{S}}_u. \tag{2.4}$$

Step 2. Data-dependent orthogonalization

For $i = 1, \ldots, n$ compute the weights to be used in Step 3:

$$\hat{f}_{uig}^2 = \Lambda'(D_{ig}\tilde{\theta}_u + X_{ig}\tilde{\beta}_u). \tag{2.5}$$

Step 3. Weighted post-lasso OLS of $[\hat{f}_u D_j]$ on $[\hat{f}_u D_{\mathcal{J}\setminus j}]$ and $[\hat{f}_u X]$

For each target variable D_j, define $\tilde{X}^j = (D_{\mathcal{J}\setminus j}, X)$ and run the weighted lasso:

$$\hat{\gamma}_u^j \in \arg\min_{\gamma} \frac{1}{2G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \hat{f}_{uig}^2 \big(D_{jig} - \tilde{X}_{ig}^j \gamma\big)^2 + \frac{\lambda_2}{G} \|\hat{\Psi}_{uj}\gamma\|_1, \tag{2.6}$$

where λ_2 is a penalty parameter and $\hat{\Psi}_{uj}$ a diagonal penalty loading matrix whose entries are chosen according to the procedure explained in Appendix A.2. Define $\hat{\mathcal{S}}_u^j = \{l \in \{\tilde{p} + 1, \ldots, \tilde{p} + p\} \mid (\hat{\gamma}_u^j)_{l-1} \neq 0\}$.

Compute the post-lasso WLS coefficients:

$$\tilde{\gamma}_u^j \in \arg\min_{\gamma} \frac{1}{2G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \hat{f}_{uig}^2 \big(D_{jig} - \tilde{X}_{ig}^j \gamma\big)^2 : \mathrm{supp}(\gamma) \subseteq \mathrm{supp}(\hat{\gamma}_u^j), \tag{2.7}$$

where $\mathrm{supp}(\hat{\gamma}_u^j)$ represents the set of indices associated with non-zero coefficients in the weighted lasso solution. Note that the post-lasso WLS coefficients are required for the computation of $\hat{\Psi}_{uj}$ and, in particular, for the construction of the influence functions defined in (2.17) since Belloni et al. (2016a) report the post-ℓ_1 estimates to possess a superior finite sample behaviour.

Step 4. Logit of Y^u on union of variables selected in either Step 1 or 3

Obtain the post-double selection logit coefficients:

$$(\check{\theta}_u, \check{\beta}_u) \in \arg\min_{\theta,\beta} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, M_u(W_{ig}, \theta, \beta) : \mathrm{supp}(\theta', \beta') \subseteq \mathcal{J} \cup \hat{\mathcal{S}}_u \bigcup_{j=1}^{\tilde{p}} \hat{\mathcal{S}}_u^j. \tag{2.8}$$

The resulting post-double selection point estimates $\check{\theta}_{u1}, \ldots, \check{\theta}_{u\tilde{p}}$ measure the impact of target regressors $D_1, \ldots, D_{\tilde{p}}$ on Y^u at threshold u.

It should be noted that our approach differs slightly from that of Belloni et al. (2018a) in that we jointly estimate all target parameters in Step 4. As pointed out in, e.g., Appendix B of Belloni et al. (2016a), we can add additional variables to $\mathrm{supp}(\theta, \beta)$ by including covariates selected in a different equation. In the present application this is a computationally attractive modification, as we avoid running separate estimations for all seven target variables over all thresholds. Moreover, this greatly facilitates the computation of the influence functions in (2.17), since we can pre-compute certain quantities based on the joint estimation in Step 4 and, thus, process the Jacobian and the outer product of the clustered orthogonal score in a vectorized manner. Generally, the post-double selection method simultaneously establishes in-sample orthogonality with respect to the instruments for all target variables by controlling for the effect of all selected confounders. Given that the cardinality of the union of selected variables is still far smaller than the sample size, we expect the inferential results obtained by a joint estimation to be more robust when compared to target-by-target computations; see, e.g. Appendix L in Belloni et al., 2018b.

2.2. *Inference*

In order to compute pointwise and simultaneous confidence intervals for sets of coefficients $(\theta_{uj}), u \in \mathcal{U}, j \in \mathcal{J}$ we use the following procedures.

The Neyman-orthogonal moment function for target parameter θ_{uj} is given by

$$\psi_j(W, \theta_u, \eta_u) = \big\{Y^u - \Lambda(D\theta_u + X\beta_u)\big\} \cdot \big(D_j - \tilde{X}^j \gamma_u^j\big), \tag{2.9}$$

where the nuisance parameters are collected in $\eta_u = (\beta_u', \gamma_u^{1\prime}, \ldots, \gamma_u^{\tilde{p}\prime})'$. Define

$$\psi(W, \theta_u, \eta_u) = (\psi_1(W, \theta_u, \eta_u), \ldots, \psi_{\tilde{p}}(W, \theta_u, \eta_u))' \tag{2.10}$$

and the Jacobian matrix

$$J(W, \theta_u, \eta_u) = \frac{\partial\, \psi(W, \theta_u, \eta_u)}{\partial\, \theta_u'}. \tag{2.11}$$

As shown in Belloni et al. (2018b), the post-double selection procedure enforces the sample moment condition

$$\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u) = 0, \tag{2.12}$$

where $\check{\eta}_u = (\check{\beta}_u', \tilde{\gamma}_u^{1\prime}, \ldots, \tilde{\gamma}_u^{\tilde{p}\prime})'$. Given the estimates for the nuisance parameters $\check{\eta}_u$, an expansion for the target parameters $\check{\theta}_{u1}, \ldots, \check{\theta}_{u\tilde{p}}$ yields a consistent estimate of their asymptotic covariance matrix

$$\hat{\Sigma}^u = \hat{J}_u^{-1} \hat{B}_u \hat{J}_u^{-1\prime} \tag{2.13}$$

with

$$\hat{J}_u^{-1} = \Big[\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, J(W_{ig}, \check{\theta}_u, \check{\eta}_u) \Big]^{-1} \tag{2.14}$$

$$\hat{B}_u = \Big[\frac{1}{G} \sum_{g=1}^{G} \Big(\sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u) \Big) \Big(\sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u) \Big)' \Big], \tag{2.15}$$

where the inner part of matrix $\hat{B}_u$ accounts for the clustering of observations. The estimated asymptotic variance of target parameter θ_{uj} is given by $\hat{\sigma}_{uj}^2 = (\hat{\Sigma}_{j,j}^u)$.

The Neyman-orthogonal moment condition for the target parameters $\check{\theta}_u = (\check{\theta}_{u1}, \ldots, \check{\theta}_{u\tilde{p}})'$ is constructed such that

$$\frac{\partial}{\partial \eta} \Big[\frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n_g} v_{ig}\, \psi(W_{ig}, \check{\theta}_u, \eta) \Big] \Big|_{\eta = \check{\eta}_u} = 0, \tag{2.16}$$

implying that the estimating equations are first-order immune with respect to the nuisance terms, i.e., by constructing an instrument for D_j in (2.9), one has “partialled-out” the effect of covariates $\tilde{X}^j$.

For the multiplier bootstrap procedure, define the $\tilde{p}$-dimensional estimated influence function as

$$\text{influence}(W_{ig}, \check{\theta}_u, \check{\eta}_u) = -\hat{J}_u^{-1} \psi(W_{ig}, \check{\theta}_u, \check{\eta}_u). \tag{2.17}$$

The multiplier bootstrap critical value c_α is computed as the $(1-\alpha)$-quantile of the distribution of

$$\hat{s} = \sup_{u \in \mathcal{U}, j \in \mathcal{J}} \frac{1}{\sqrt{G} \hat{\sigma}_{uj}} \sum_{g=1}^{G} \xi_g \cdot \Big[\sum_{i=1}^{n_g} v_{ig} \cdot \text{influence}_j(W_{ig}, \check{\theta}_u, \check{\eta}_u) \Big],\ \xi_g \sim \text{iid}\ \mathcal{N}(0,1) \tag{2.18}$$

where $\text{influence}_j(W_{ig}, \check{\theta}_u, \check{\eta}_u)$ is the j-th element of $\text{influence}(W_{ig}, \check{\theta}_u, \check{\eta}_u)$. The distribution of $\hat{s}$ can be obtained by repeatedly drawing weights ξ_g from the standard normal distribution.

The bootstrap critical value c_α is then used as a scaling factor for the pointwise confidence regions which results in a simultaneous confidence band covering multiple target parameters $(\theta_{uj}), j \in \mathcal{J}$ at multiple distribution thresholds $u \in \mathcal{U}$ with probability $1-\alpha$, i.e.,

$$\mathrm{P}\Big(\check{\theta}_{uj} - c_\alpha \frac{\hat{\sigma}_{uj}}{\sqrt{G}} \;\leq\; \theta^0_{uj} \;\leq\; \check{\theta}_{uj} + c_\alpha \frac{\hat{\sigma}_{uj}}{\sqrt{G}} \;\forall\; u \in \mathcal{U}, j \in \mathcal{J}\Big) \approx 1 - \alpha. \tag{2.19}$$

3. DATA AND IMPLEMENTATION

3.1. Data sources and specification

Our empirical analysis is based on the German Socio-Economic Panel Study (SOEP, v35) which is a long-running survey providing representative information about the German population; see, e.g. Schröder et al. (2020). We utilize information from 2011 to 2018, which encompasses the years leading up to and following the implementation of the German minimum wage on January 1, 2015. The strength of a survey like the SOEP is the wealth of information that can be used as covariates. A weakness is the moderate sample size of around 11,000 wage earners per year, which motivates the use of specification selection methods. After applying selection criteria, our final sample includes approximately 90,000 observations. We rely only on cross-sectional information in the SOEP, as frequent refreshment samples and permanent dropout make the panel highly unbalanced. However, we fully account for longitudinal correlation in our data by adjusting our inference and penalty selection procedures as described in section 3.3. In addition, we use the sampling weights provided with the SOEP which ensure that cross-sectional information is representative for the German population in the given year. We exclude from our sample individuals who are not subject to the minimum wage (the self-employed, students, apprentices, interns and similar groups).

Following the seminal work by Card (1992) on minimum wages, we assess the effects of the minimum wage introduction using a continuous treatment indicator, called the minimum wage bite MWB_{it}, which represents the share of workers in specific population subgroups who earned less than 8.50 euros/hour before the policy was implemented. The minimum wage is expected to have its most significant impact on subgroups with the highest pre-reform exposure, when other relevant factors are controlled for.[3] As the sample size of the SOEP would be too low to construct reliable bite measures for small population subgroups, we take our bite measure from a larger data set, the German Structure of Earnings Survey (GSES). The bite measure used here is defined at the 2-digit industry level differentiated by East/West Germany. Defining the bite measure at the industry level is well aligned with the structure of industrial relations in Germany, where a substantial part of wage bargaining takes place at the industry level. Our bite measure varies between .003 and .701, providing large variation to measure the effects induced by the exposure to the newly introduced minimum wage.[4]

Note that MWB_{it} is indexed in both i and t because the minimum wage bite, measured at the industry level in 2014, is assigned annually based on each individual's industry

[3] We thank a reviewer for suggesting a robustness check adjusting the minimum wage bite for counterfactual wage growth due to inflation or productivity gains. However, this adjustment is unlikely to affect our results, as inflation was near zero at the time of the minimum wage introduction and productivity growth at the lower end of the wage distribution is negligible.

[4] For more details on the bite measure used here, see Biewen et al. (2022). An alternative would be to define the bite at the level of labour market regions but this faces the difficulty that the coverage of labour market regions in a survey like the SOEP is patchy and that regional information in the SOEP can only be processed on-site with limited computational facilities.

affiliation. Consequently, our estimations also capture wage effects for individuals who switch industries.

We measure the effects of the minimum wage introduction on the distribution of our outcome variables $Y \in \{$"hourly wage", "hours worked", "monthly earnings"$\}$ by the following difference-in-differences specification:

$$\begin{aligned}\mathrm{P}(Y_{it} < u \mid D_{it}, X_{it}) = \Lambda(&\theta_{u,bite} \cdot MWB_{it} + \theta_{u,2011/12} \cdot \mathbf{1}\{t = 2011/12\} \\ &+ \theta_{u,2015/16} \cdot \mathbf{1}\{t = 2015/16\} + \theta_{u,2017/18} \cdot \mathbf{1}\{t = 2017/18\} \\ &+ \Theta_{u,2011/12} \cdot MWB_{it} \cdot \mathbf{1}\{t = 2011/12\} \\ &+ \Theta_{u,2015/16} \cdot MWB_{it} \cdot \mathbf{1}\{t = 2015/16\} \\ &+ \Theta_{u,2017/18} \cdot MWB_{it} \cdot \mathbf{1}\{t = 2017/18\} + X_{it}\beta_u) \end{aligned} \tag{3.1}$$

for $u \in \mathcal{U}$. In order to keep the number of target coefficients low, we combine two adjacent years. The relevant periods are: $t = 2013/14$ representing the reference period, i.e. the period immediately before the minimum wage introduction, $t = 2015/16$ representing short-term effects after the introduction, $t = 2017/18$ representing medium-term effects after the introduction, and $t = 2011/12$ representing the pre-test period.

Coefficients $\Theta_{u,2015/16}$ and $\Theta_{u,2017/18}$ correspond to the short-term and medium-term treatment effects of the minimum wage introduction on the likelihood of falling below a particular threshold u in the outcome distribution. They measure to what extent, e.g., hourly wages below a particular level u became more or less frequent after the minimum wage introduction per unit of exposure to the newly introduced minimum wage MWB_{it}, controlling for time effects $\mathbf{1}\{t = 2011/12\}$, $\mathbf{1}\{t = 2015/16\}$, $\mathbf{1}\{t = 2017/18\}$, base effects MWB_{it}, and for all other characteristics X_{it} that are relevant for explaining that a particular wage observation falls below threshold u (work experience, education, occupational characteristics, see below). Coefficient $\Theta_{u,2011/12}$ provides a pre-treatment test as it measures to what extent differences already emerged between high and low exposure groups in the pre-treatment period, i.e. between 2011/12 and the reference period 2013/14. The main (uninteracted) effects in the difference-in-differences specification control for general time differences $\theta_{u,2011/12}$, $\theta_{u,2015/16}$, $\theta_{u,2017/18}$, and for differences $\theta_{u,bite}$ between high and low bite groups of falling below threshold u that are time-invariant. All other terms X_{it} of the regression are selected by the lasso procedure.

3.2. Variables and feature engineering

Our dependent variables are derived from the survey information on monthly earnings and actual hours worked per week (including overtime). Monthly earnings and actual hours worked per week are taken as they appear in the survey. Hourly wages are computed as monthly earnings divided by monthly hours worked (defined as weekly hours times 4.345).

Table B.1 in the Appendix describes the information that is used to construct the set of control candidates (features) from which the ℓ_1-methods can choose relevant elements for predicting the nuisance terms at each threshold. The total number of features constructed in this way is in the order of several thousands, as we not only include transformations of continuous variables (e.g. polynomial terms, square root, log) and indicators for potentially important individual values of continuous variables (e.g. an indicator for having an unemployment experience of zero years), but also interactions and full sets of indicators for all our categorial variables. To illustrate this, consider an educational classification with five categories. In this case, we include a full set of five indicators describing the membership in each category (no omitted category). The lasso can then flexibly pick the indicators that help to remove the omitted variable bias for explaining the effect of the treatment variables at a particular threshold.[5] It is important not to omit a reference category when constructing sets of such indicators as exactly the omitted category could be the one preferred by the lasso (the information represented by the omitted category could be re-constructed as a linear combination of other categories, but this runs counter to the idea of finding a sparse approximation for the nuisance term). In a similar manner, we offer the lasso nested or overlapping information from classifications of higher or lower aggregation levels, from which it can select the information that is most suitable for the purpose of removing omitted variable bias. For example, we incorporate occupation codes at varying aggregation levels (1-digit, 2-digit, etc.) and nested or partly overlapping education classifications that provide both finer and more coarse information (see Table B.1).

In order to arrive at the final set of potential covariates offered to the lasso, we eliminated from the full set of features described in Table B.1 i) constant features, ii) duplicates/multiples of other features, iii) features that uniquely characterize less than 1 percent of our sample. We could relax restriction iii) to a certain extent without changing results in important ways. However, our experience suggests that doing so can result in an increased likelihood of perfect prediction problems and convergence issues in the logit models. This is in contrast to the primary motivation of this paper, which is to identify a fully automatic method for selecting controls at the typically large number of thresholds without the need to manually fix problems or eliminate features at individual thresholds. Applying the above criteria, the final number of features included in our estimations was around 2,500 (the exact number of features depends on the outcome variable as features related to working hours cannot be included for monthly earnings and hours worked due to perfect prediction issues). This feature set is clearly too large for individual specification searches at one given threshold, let alone at the typically around 40-50 thresholds used per dependent variable in our application.

3.3. Details on lasso implementation

As described in section 2, our methods allow for clustering of observations in two ways: i) for statistical inference and ii) for the choice of lasso penalties. With regard to i), it is well-known that in difference-in-differences-like designs, it is necessary to cluster at the level of the treatment variable; see, e.g. Abadie et al. (2022). Our treatment variable MWB_{it} is based on the combination of 2-digit industries and East/West information. This provides 152 population subgroups at the level of which we cluster in all inference procedures (estimation of variance matrices and draws for multiplier bootstrap). We initially also tried to cluster the lasso penalties at this level but found that this led to quite erratic and volatile results across different thresholds and coefficients which did not seem plausible. This behaviour of the lasso is not surprising given the relatively low number of clusters in our application and their sometimes chunky nature. For computing

[5] For categorial variables, we also define a category “missing value” that may also be picked by the lasso if it helps to predict the treatment or the outcome variable. This also helps to conserve the number of observations as observations with missings in these variables do not have to be discarded.

lasso penalty loadings, we therefore clustered at the level of the panel units, which is standard for panel data; see, e.g. Belloni et al. (2016b) and Ahrens et al. (2020).

As mentioned above, it is useful to offer potentially multi-collinear control variables to the lasso (nested information, full sets of indicators etc.) to allow the lasso to extract the information that is most suitable for removing omitted variable bias. It is well known that the lasso solution need not be unique if the feature set contains mostly discrete (binary) variables as in our case; see Tibshirani (2013). To evade numerical difficulties, we implemented our post-ℓ_1-estimators using the Moore-Penrose pseudo-inverse.

In our application, the cardinality of the active set of control variables was between 100 and 120 depending on the threshold. The double-selected features consistently included information on educational qualifications, work experience as well as additional controls that differed across thresholds in plausible ways (e.g., indicators for low occupational positions/job types at lower thresholds, information on firm characteristics or particular educational/occupational qualifications at medium or upper thresholds, interactions of such characteristics with gender or East/West Germany at particular thresholds).

4. EMPIRICAL RESULTS AND ECONOMETRIC ANALYSIS

We begin by examining how the minimum wage affected the likelihood that hourly wages fall below certain thresholds. Figures 1abc display the treatment effect coefficients for the pre-treatment period, $\Theta_{u,2011/12}$, and for the two evaluation periods, $\Theta_{u,2015/16}$ and $\Theta_{u,2017/18}$. The results show that, as intended by policymakers, the likelihood of having hourly wages below 8.50 euros/hour declined in groups with high minimum wage exposure after the introduction (Figures 1bc). In the pre-test period, there were no significant differences between treated and untreated individuals (Figure 1a). In 2017/18, there is a slight indication of spill-over effects beyond the threshold of 8.50 euros/hour. However, these effects are not statistically significant (see Figure 1c).[6]

Indeed, the pattern in the lower part of the distribution in Figures 1bc might also reflect disemployment effects, since our analysis includes only employed workers. In separate analyses, however, Caliendo et al. (2019), Dustmann et al. (2022), Bossler and Schank (2023) have found little evidence for such disemployment effects. Link (2024) concludes that firms affected by the minimum wage mostly increased prices but did not cut employment. Below, we also show that there were no minimum wage effects on low levels of working hours, which is what one would expect if the minimum wage systematically displaced low-wage workers.

To assess the economic magnitude of the effects in Figures 1abc, note that logit coefficients indicate changes in the log-odds ratio. For a worker i in year t with an initial probability p_{it} of falling below a wage threshold u, an increase of Δ_{MWB} in treatment intensity changes the log-odds ratio as follows:

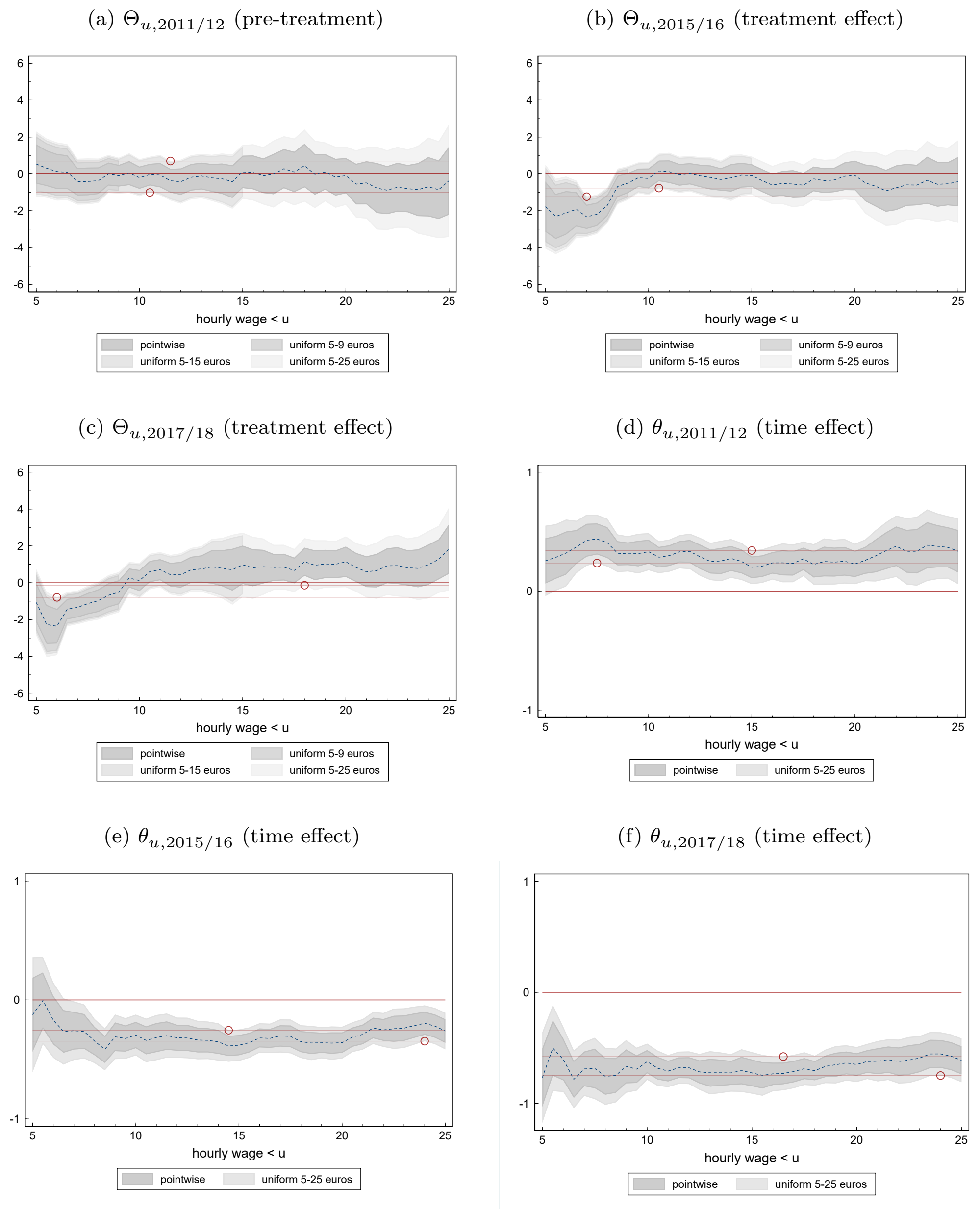
$$\log \frac{p'_{it}}{1-p'_{it}} - \log \frac{p_{it}}{1-p_{it}} = \Theta_u \cdot \Delta_{MWB}, \tag{4.1}$$

where Θ_u is the treatment effect logit coefficient, and p'_{it} is the updated probability of falling below a threshold u. This updated probability p'_{it} is computed as:

$$p'_{it} = \Lambda(m_{it}), \text{ where } m_{it} = \log \frac{p_{it}}{1-p_{it}} + \Theta_u \cdot \Delta_{MWB}. \tag{4.2}$$

[6]Note that the minimum wage level was initially set to 8.50 euros/hour in 2015, but increased to 8.84 euros/hour in 2017.

Figure 1: Coefficient processes for hourly wage distribution



(a) $\Theta_{u,2011/12}$ (pre-treatment)

(b) $\Theta_{u,2015/16}$ (treatment effect)

(c) $\Theta_{u,2017/18}$ (treatment effect)

(d) $\theta_{u,2011/12}$ (time effect)

(e) $\theta_{u,2015/16}$ (time effect)

(f) $\theta_{u,2017/18}$ (time effect)

Note: These graphs show the processes of estimated logit coefficients indexed by thresholds $u \in \mathcal{U}$. Grey shaded areas represent the 90% uniform confidence bands based on 100,000 multiplier bootstrap replications. The intervals in (a)-(c) refer to increasing intervals of distributional thresholds. The circles mark the infimum of the upper confidence bounds and the supremum of the lower confidence bounds.

As an example, consider the logit coefficients of around minus two for wage thresholds below 8.50 euros in Figure 1b. For a worker i in year t, whose initial probability of falling below a threshold is $p_{it} = 0.7$, an increase in the exposure to the minimum wage of $\Delta_{MWB} \in \{0.1, 0.2, 0.3\}$ reduces the probability of falling below a threshold to $p'_i \in \{0.656, 0.610, 0.561\}$. These are substantial reductions.

Figure 1 presents simultaneous confidence intervals, computed using the multiplier bootstrap over an increasing range of thresholds. By design, these bands widen as more coefficients are included, most notably when shifting from pointwise to simultaneous intervals. The graphs also show a solid red line representing a zero effect. If there is at least one threshold, at which the simultaneous confidence band does not include this zero line, we reject the hypothesis that the minimum wage had no effect on the distribution of hourly wages. This hypothesis is not rejected in the pre-test period 2011/12 (Figure 1a), but rejected in the post-introduction periods 2015/16 and 2017/18 (Figures 1bc). Analogously, we can also test whether effects are homogeneous across the distribution. To do so, the infimum of the upper confidence bounds has to be compared to the supremum of the lower confidence bounds. In the graphs, these points are symbolized by small circles. If the supremum circle lies above the infimum circle, the hypothesis of a constant effect over the entire distribution is rejected. Such effect homogeneity can be rejected for both post-introduction periods 2015/16 (Figure 1b) 2017/18 (Figure 1c), but not for the pre-test period 2011/12 (Figure 1a).

Figure 2: Coefficient process for hourly wage distribution: base effect

(a) $\theta_{u,bite}$ (bite indicator)

Note: This graph shows the process of estimated logit coefficients indexed by thresholds $u \in \mathcal{U}$. Grey shaded areas represent the 90% uniform confidence bands based on 100,000 multiplier bootstrap replications. The circles mark the infimum of the upper confidence bounds and the supremum of the lower confidence bounds.

In order to illustrate the benefits of our method, we also show the results for the main effects in the difference-in-differences specification (time effects and base effect). The time effects displayed in Figures 1def indicate uniform wage growth over time: wages were uniformly lower in the pre-test period 2011/12 compared to the reference period 2013/14 (Figure 1d), higher in the first post-reform period 2015/16 (Figure 1e) and even more so in the second post-reform period 2017/18 (Figure 1c).[7] There is an indication of less wage growth at the lower end compared to the rest of the distribution in 2015/16, but homogeneity cannot be rejected. The base effect of the difference-in-differences specification displayed in Figure 2 shows that wages were predominantly lower in high-bite groups, both preceding and succeeding the implementation of the reform. This is what one would expect because wages in high-bite industries are likely to be generally lower.

Figures 3abc examine how the minimum wage affected the distribution of monthly earnings, shedding light on which segments of the personal income distribution benefited most. The simultaneous confidence intervals in these figures extend up to 2,050 euros/month, since it is implausible that minimum wage recipients earn more than this amount. In fact, earning over 2,050 euros/month at the minimum wage rate would require working more than 50 hours a week. Our results suggest that the largest gains occurred not among individuals with very low monthly wages, i.e., marginal part-time workers, but among those earning between 700 and 1,700 euros/month, as indicated by the negative coefficients in the post-treatment periods. It is important to note, however, that while the coefficients turn negative in the post-treatment periods, and the hypothesis of no effect can thus be rejected, the uniform confidence intervals are wide, rendering these estimates only marginally significant. This is the likely reason why previous contributions based on the same data set struggled to establish significant effects of the minimum wage introduction on monthly earnings; see Burauel et al. (2019, 2020) and Caliendo et al. (2022).

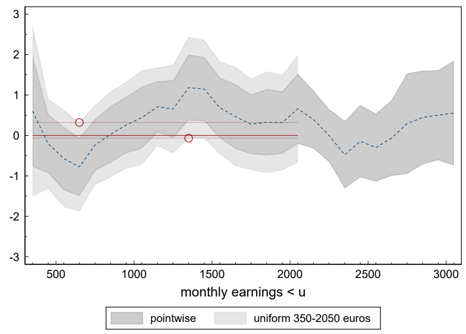
To further illustrate the usefulness of the framework proposed by Belloni et al. (2018a), we demonstrate in Figures 3bc the point made by Roth (2022) that conditioning on the result of a pre-test leads to an under-coverage of confidence intervals for difference-in-differences effects. In accordance with equation (2.19), confidence intervals pertaining to the process of a single target coefficient over arbitrary sets of distributional thresholds, as well as the processes of multiple target coefficients, can be constructed. In Figures 3bc, we incorporate the coefficients for potential pre-trends from Figure 3a into the simultaneous interval. This widens confidence intervals and demonstrates the "cost" in interval coverage for the treatment effect coefficients associated with testing for the existence of pre-trends. Although this cost is small in our application, it is non-negligible and further reduces the significance of the treatment effect estimates.

An open question arising from Figures 1abc and 3abc is the extent to which the observed difference in the effects of the minimum wage on monthly and hourly earnings can be explained by potential changes in working hours, as suggested by Burauel et al. (2019, 2020) and Caliendo et al. (2022). The results depicted in Figures 3def provide no evidence of significant changes in the distribution of working hours as a result of the minimum wage. This is particularly evident in the second post-reform period 2017/18. Here, we observe a uniform zero effect, as the zero line is fully encompassed in both the pointwise and uniform confidence intervals.

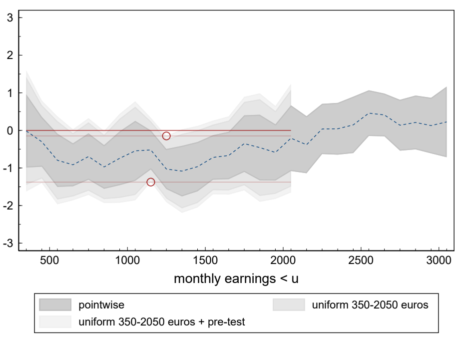
[7] Note that positive/negative coefficients indicate a higher/lower likelihood of falling below a particular wage threshold, i.e. lower/higher wages.

Figure 3: Coefficient processes for earnings and working hours distributions

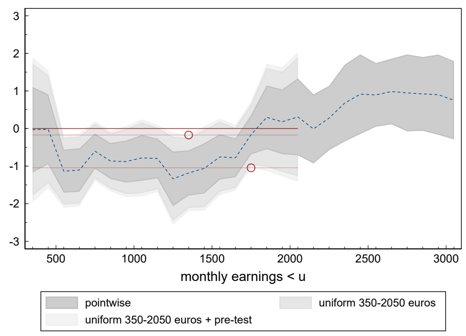
(a) $\Theta_{u,2011/12}$ (earnings)



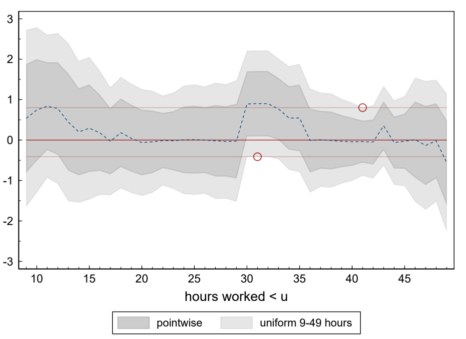
(b) $\Theta_{u,2015/16}$ (earnings)



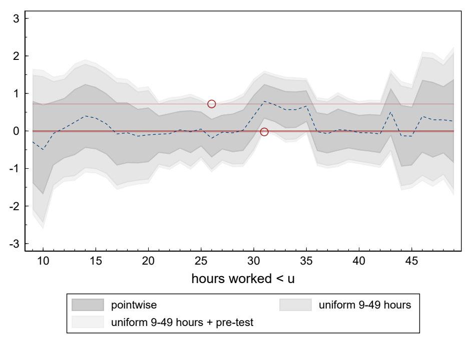
(c) $\Theta_{u,2017/18}$ (earnings)



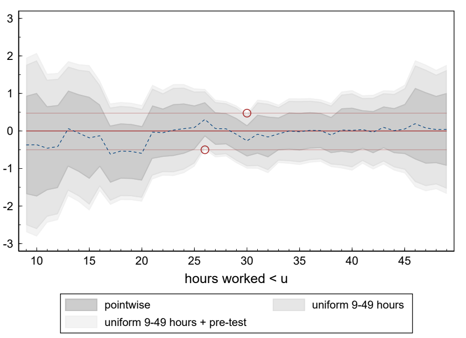
(d) $\Theta_{u,2011/12}$ (working hours)



(e) $\Theta_{u,2015/16}$ (working hours)



(f) $\Theta_{u,2017/18}$ (working hours)



Note: These graphs show the processes of estimated logit coefficients indexed by thresholds $u \in \mathcal{U}$. Grey shaded areas represent the 90% uniform confidence bands based on 100,000 multiplier bootstrap replications. The intervals in (a)-(c) refer to the monthly earnings range relevant to minimum wage recipients (350 to 2050 euros). The additional intervals in (b) and (c) simultaneously cover the coefficient process of the respective period and the coefficients of the pre-test period 2011/12. The circles mark the infimum of the upper confidence bounds and the supremum of the lower confidence bounds.

In summary, our results indicate that the introduction of the minimum wage neither led to significant reductions in weekly working hours, which would help maintain monthly wage bills at a constant level, nor triggered substantial transitions between employment categories (part-time, marginal part-time, and full-time).

5. CONCLUSION

This paper uses a distribution regression model to evaluate the effects of the introduction of the German minimum wage in 2015 on the distribution of hourly wages, hours worked and monthly earnings. Our data source is the German Socio-Economic Panel (SOEP) which is characterized by a moderate sample size but a large number of potential control variables. We measure the effects of the minimum wage at each point of our outcome distributions employing flexible machine learning methods recently developed by Belloni et al. (2018a). These methods allow us to automatically specify a large number of parallel logit models over a fine grid of distributional thresholds, while providing valid statistical inference across ranges of thresholds, after a thorough, machine-led specification search. Our distribution regression analysis provides a more comprehensive picture about the points of the distribution at which the minimum wage had an effect compared to previous contributions. It also allows us to assess the information content of our data base in an objective manner, unaffected by subjective decisions (p-hacking) or pre-tested specification choices. Our findings suggest that the minimum wage displaced hourly wages below its minimum level, benefited monthly wages in the lower-middle segment of the distribution but not at the lowest end, and did not significantly alter the distribution of working hours. These findings help reconcile conflicting results in the literature regarding the effects of the German minimum wage, which had previously varied depending on the data source used.

ACKNOWLEDGEMENTS

We thank the editor, Petra Todd and two anonymous reviewers whose comments helped us improving the initial version of the paper. We also thank Marco Caliendo, Bernd Fitzenberger, Martin Huber, Michael Knaus, Marian Rümmele, Uta Schönberg, Martin Spindler as well as participants at EALE 2023 and at the "Machine Learning Meets Quantitative Social Science" Workshop at the University of Tübingen for many valuable comments and discussions. Further thanks go to Nicolas Erhardt for excellent research assistance. This paper uses data from the Socio-Economic Panel study SOEP-Core, v35, doi:10.5684/soep-core.v35. Martin Biewen is a member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2064/1 - Project 390727645.

REFERENCES

Abadie, A., S. Athey, G. Imbens and J. Wooldridge (2022). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, *138*, 1–35.

Ahrens, A., C.B. Hansen, A.B. Schaffer (2020). lassopack – model selection and prediction with regularized regression in Stata. *Stata Journal*, *20*, 176–235.

Belloni, A., V. Chernozhukov and Y. Wei (2016a). Post-selection inference for generalized linear models with many controls. *Journal of Business and Economic Statistics*, *34*, 606–619.

Belloni, A., V. Chernozhukov, C. Hansen and D. Kozbur (2016b). Inference in high-dimensional panel models with an application to gun control. *Journal of Business and Economic Statistics*, *34*, 590–605.

Belloni, A., V. Chernozhukov, D. Chetverikov and Y. Wei (2018a). Uniformly valid post-regularization confidence regions for many functional parameters in a z-estimation framework. *Annals of Statistics*, *46*, 3643–3675.

Belloni, A., V. Chernozhukov, D. Chetverikov and Y. Wei (2018b). *Supplement to "Uniformly valid post-regularization confidence regions for many functional parameters in a z-estimation framework"*.

Biewen, M., B. Fitzenberger and M. Rümmele (2022). Using distribution regression difference-in-differences to evaluate the effects of a minimum wage introduction on the distribution of hourly wages and hours worked. *IZA Discussion Paper No. 15534*, IZA Institute of Labor Economics, Bonn.

Bossler, M. and T. Schank (2023). Wage inequality in Germany after the minimum wage introduction. *Journal of Labor Economics*, *41*, 813–855.

Burauel, P., M. Caliendo, M. Grabka, C. Obs, M. Preuss and C. Schröder (2019). The impact of the German minimum wage on individual wages and monthly earnings. *Journal of Economics and Statistics*, *240*, 201–231.

Burauel, P., M. Caliendo, M. Grabka, C. Obs, M. Preuss and C. Schröder (2020). The impact of the minimum wage on working hours. *Journal of Economics and Statistics*, *240*, 233–267.

Caliendo, M., C. Schröder and L. Wittbrodt (2019). The causal effects of the minimum wage introduction in Germany – an overview. *German Economic Review*, *20*, 257–292.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder and L. Wittbrodt (2022). The short- and medium-term distributional effects of the Germany minimum wage reform. *Empirical Economics*, *64*, 1149–1175.

Card, D. (1992). Using regional variation in wages to measure the effects of the federal minimum wage. *Industrial Labor Relations Review*, *46*, 22–37.

Chernozhukov, V., I. Fernandez-Val, B. Melly (2013). Inference on counterfactual distributions. *Econometrica*, *81*, 2205–2268.

Chiang, H.D. (2020) *Three essays in cluster robust machine learning and high-dimensional econometrics.* Ph.D. dissertation, Graduate School Vanderbilt University.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, P. vom Berge (2022). Reallocation effects of the minimum wage. *Quarterly Journal of Economics*, *137*, 267–328.

Foresi, S. and F. Peracchi (1995). The conditional distribution of excess returns: an empirical analysis. *Journal of the American Statistical Association*, *90*, 451–466.

Link, S. (2024). The price and employment response of firms to the introduction of minimum wages. *Journal of Public Economics*, *239*, 1–18.

Mindestlohnkommission (2020). *Dritter Bericht zu den Auswirkungen des gesetzlichen Mindestlohns.* Mindestlohnkommission, Bundesamt für Arbeitsschutz und Arbeitsmedizin (BAuA), Berlin.

Roth, J. (2022). Pretest with caution: event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*, 305–322.

Schröder, C., J. König, A. Fedorets, J. Goebel, M. Grabka, L. Lüthen, M. Metzing, F. Schikora, S. Liebig (2020). The economic research potential of the German Socio-Economic Panel study. *German Economic Review*, *31*, 335–371.

Tibshirani, R.J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics*, *7*, 1456–1490.

APPENDIX A: CHOICE OF PENALTY LEVELS AND LOADINGS

In order to allow for a different choice of observation clusters when computing penalties as opposed to computing standard errors and confidence intervals, we define in the present section clusters as $g^* = 1, \dots, G^*$ with observations $i = 1, \dots, n_{g^*}$ (as opposed to the main text, where we defined clusters as $g = 1, \dots, G$ with observations $i = 1, \dots, n_g$). In addition, define the $(\tilde{p} + p)$-dimensional vector $\tilde{X} = (D, X)$ and the $((\tilde{p} - 1) + p)$-dimensional vector $\tilde{X}^j = (D_{\mathcal{J} \setminus j}, X)$ for $j \in \mathcal{J} = \{1, \dots, \tilde{p}\}$. The algorithms for setting penalty levels and loadings are due to Belloni et al. (2018a) with modifications for clustering of observations as in Chiang (2020). In addition, we incorporate sampling weights v_{ig}.

A.1. Penalty level and loadings for logistic lasso

STEP 1. Initialize procedure

Define some tolerance level, e.g., $\epsilon = 10^{-4}$, $\bar{m}$ as the maximal number of iterations and set the penalty level to $\lambda_1 = c\sqrt{G^*}\,\Phi^{-1}(1 - \gamma/2(\tilde{p} + p))$, where $c = 1.1$ and $\gamma = 0.1/\log(G^*)$.
Initialize $m = 0$. Starting from a constant-only model, determine the five features that have the maximal *ex-ante* gradient by absolute value. These represent the five most promising features for reducing the prediction error. Fill their post-ℓ_1 logit coefficients into $(\tilde{\theta}_u^{m\prime}, \tilde{\beta}_u^{m\prime})'$ and set all other entries equal to zero.

STEP 2. Iterative determination of penalty and loadings

Update the diagonal elements of $\hat{\Psi}_u$ according to:

$$\hat{l}_{uk}^{m+1} = \left\{ \frac{1}{G^*} \sum_{g^*=1}^{G^*} \Big[\sum_{i=1}^{n_{g^*}} v_{ig^*} \, (Y_{ig^*}^u - \Lambda(D_{ig^*}\tilde{\theta}_u^m + X_{ig^*}\tilde{\beta}_u^m))\tilde{X}_{kig^*} \Big]^2 \right\}^{1/2},$$

$$\hat{\Psi}_u^{m+1} = \mathrm{diag}\Big(\hat{l}_{uk}^{m+1},\ k = 1, \dots, (\tilde{p} + p)\Big).$$

Run post-ℓ_1 logistic regression to obtain refined logit coefficients $(\tilde{\theta}_u^{m+1\prime}, \tilde{\beta}_u^{m+1\prime})'$.
If $\max_{1 \le k \le (\tilde{p}+p)} |\hat{l}_{uk}^{m+1} - \hat{l}_{uk}^m| < \epsilon$ or $m = \bar{m}$ then stop. Otherwise set $m \leftarrow m + 1$ and repeat Step 2.

A.2. Penalty level and loadings for WLS lasso

STEP 1. Initialize procedure

Define some tolerance level, e.g., $\epsilon = 10^{-4}$, $\bar{m}$ as the maximal number of iterations and set the penalty level to $\lambda_2 = c\sqrt{G^*}\,\Phi^{-1}(1 - \gamma/2(\tilde{p}+p)(\tilde{p}+p-1))$, where $c = 1.1$ and $\gamma = 0.1/\log(G^*)$.
Initialize $m = 0$. Starting from a constant-only model, determine the five features that have the maximal *ex-ante* gradient by absolute value. These represent the five

most promising features for reducing the prediction error. Fill their WLS coefficients into $\tilde{\gamma}_u^{j,m}$ and set all other entries equal to zero.

Step 2. Iterative determination of penalty and loadings

Update the diagonal elements of $\hat{\Psi}_{uj}$ according to:

$$\hat{l}_{ujk}^{m+1} = \left\{ \frac{1}{G^*} \sum_{g^*=1}^{G^*} \left[\sum_{i=1}^{n_{g^*}} v_{ig^*} \cdot \hat{f}_{uig^*}^2 \cdot (D_{jig^*} - \tilde{X}_{ig^*}^j \tilde{\gamma}_u^{j,m}) \tilde{X}_{kig^*}^j \right]^2 \right\}^{1/2},$$

$$\hat{\Psi}_{uj}^{m+1} = \text{diag}\left(\hat{l}_{ujk}^{m+1},\ k = 1, \ldots, (\tilde{p} - 1) + p \right).$$

Run post-ℓ_1 WLS to obtain refined WLS coefficients $\tilde{\gamma}_u^{j,m+1}$.
If $\max_{1 \leq k \leq (\tilde{p}-1+p)} |\hat{l}_{ujk}^{m+1} - \hat{l}_{ujk}^{m}| < \epsilon$ or $m = \bar{m}$ then stop. Otherwise set $m \leftarrow m + 1$ and repeat Step 2.

APPENDIX B: CONSTRUCTION OF LASSO FEATURE SET

Table B.1: Variables and transformations included in the post-double selection algorithm.

Variable	Type	Transformations included
Worker characteristics 1		
Gender	categorial(2)	indicators for each category
East/West Germany	categorial(2)	indicators for each category
Worker characteristics 2		
Age	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Years of education	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Full-time experience (years)	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Part-time experience (years)	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Full-time + 0.5 Part-time experience	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Tenure (years)	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·)
Overtime (hours/week)	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·), indicator for no overtime
Unemployment experience (years)	continuous	4-th order polynomial, $\sqrt{\cdot}$, log(·) indicator for no unemployment experience
Worker characteristics 3		
Type of school degree	categorial(9)	indicators for each category
Type of vocational training degree	categorial(7)	indicators for each category
Type of tertiary degree	categorial(11)	indicators for each category
Fine type of tertiary degree	categorial(23)	indicators for each category
Variants of no educational degree	categorial(4)	indicators for each category
ISCED classification of educational degree	categorial(10)	indicators for each category
5-group categorization German education system	categorial(5)	indicators for each category
3-group categorization German education system	categorial(3)	indicators for each category

Continued on following page

Table B.1, continued

Variable	**Type**	**Transformations included**
ISCO08 occupation code (2-digit)	categorial(40)	indicators for each category
ISCO08 occupation code (3-digit)	categorial(121)	indicators for each category
KldB2010 occupation code (1-digit)	categorial(10)	indicators for each category
KldB2010 occupation code (2-digit)	categorial(37)	indicators for each category
Occupational position	categorial(12)	indicators for each category
NACE industry code (1-digit)	categorial(18)	indicators for each category
NACE industry code (2-digit)	categorial(86)	indicators for each category
Full-time/part-time/marginal part-time	categorial(3)	indicators for each category indicator for part-time/marginal part-time combined
Minjob contract	categorial(2)	indicators for each category
Firm size categorization I (coarse)	categorial(5)	indicators for each category
Firm size categorization II (finer)	categorial(8)	indicators for each category
Public sector	categorial(3)	indicators for each category
Federal state	categorial(16)	indicators for each category
Urban area	categorial(2)	indicators for each category
Nationality (continents)	categorial(5)	indicators for each category
Nationality (subcontinents)	categorial(12)	indicators for each category
Nationality (countries)	categorial(116)	indicators for each category indicator for German nationality
Household size	count(16)	1st power, indicator for each category
Partner lives in household	categorial(3)	indicators for each category
Marital status (single/divorced/widowed etc.)	categorial(9)	indicators for each category
Number of children in household aged 0-2 years	count	1st power, indicator for zero
Number of children in household aged 3-5 years	count	1st power, indicator for zero
Number of children in household aged 6-11 years	count	1st power, indicator for zero
Number of children in household aged 12-17 years	count	1st power, indicator for zero
Person in need of care lives in household	categorial(3)	indicators for each category
Homeowner/renter (with subcategories)	categorial(5)	indicators for each category
Health indicator	ordinal(6)	1st power, indicator for highest two values indicator for lowest two values
Interactions		
Age × Household size	full expansion of age transformations with household size	
Age × (Worker characteristics 2)	full expansion of features with age transformations	
Household size × (Worker characteristics 2)	full expansion of features with household size	
Gender × (Worker characteristics 2)	full expansion of features with gender indicators	
Gender × (Worker characteristics 3)	full expansion of features with gender indicators	
East/West × (Worker characteristics 2)	full expansion of features with East/West indicators	
East/West × (Worker characteristics 3)	full expansion of features with East/West indicators	

Note: This table summarizes and categorizes all variables, their transformations and interactions that were used to construct the set of potential controls. For categorial variables, the number of categories is reported in brackets.